

REMARKS

Entry of this response is proper under 37 CFR §1.116 since no new claims or issues are presented.

Claims 1-3, 6-12, 14-19, and 21-23 are all of the claims currently pending.

It is noted that the claim amendments are made only for more particularly pointing out the invention, and not for distinguishing the invention over the prior art, narrowing the claims or for any statutory requirements of patentability. Further, Applicant specifically states that no amendment to any claim herein should be construed as a disclaimer of any interest in or right to an equivalent of any element or feature of the amended claim.

All pending claims stand rejected under 35 USC §101 as allegedly directed to non-statutory subject matter.

Claims 1-3, 6-12, and 14-18 stand rejected under 35 USC §102(a) as allegedly anticipated by Vinod Valsalam et al., “A Framework for High-Performance Matrix Multiplication Based on Hierarchical Abstractions, Algorithms and Optimized Low-Level Kernels”, hereinafter referred to as “Vinod”.

Claim 19 stands rejected under 35 USC §103(a) as allegedly unpatentable over Vinod, further in view of Philip Alpatov et al., “PLAPACK: Parallel Linear Algebra Package Design Overview”, hereinafter referred to as “Philip.”

Claim 21-23 stand rejected under 35 USC §103(a) as allegedly unpatentable over Vinod, further in view of US Patent 5,099,447 to Myszewski.

These rejections are respectfully traversed in the following discussion.

I. THE CLAIMED INVENTION

The claimed invention is directed to a method of improving at least one of speed and efficiency when executing a linear algebra subroutine on a computer having a memory hierarchical structure including at least one cache. The method includes determining, based on sizes, for a level 3 matrix multiplication processing, which matrix will have data for a submatrix block residing in a lower level cache of said computer and which two matrices will have data for submatrix blocks residing in at least one higher level cache or a memory. Data from the selected two matrices is streamed, for executing said level 3 matrix multiplication processing, so that the submatrix block residing in said lower level cache remains resident in said lower level cache.

The present inventors have recognized that conventional linear algebra processing based on LAPACK subroutines, for example, are not optimal.

The claimed invention, on the other hand, along with various other techniques described in the co-pending applications, provides techniques that improve processing efficiency. More specifically, the present invention provides a memory management method allowing for a streaming of data through a cache, using another operand as having the “matrix role” and being resident in the cache.

II. THE STATUTORY SUBJECT MATTER REJECTION

The Examiner continues to reject all claims as allegedly directed to non-statutory subject matter.

Applicants respectfully disagree.

On page 2 of the Office Action, the Examiner’s states:

“Claims 1-3, 6-12, 14-19, and 21-23 cite a method, apparatus, and medium for performing matrix multiplication in [a] computer in accordance with a mathematical algorithm. However, claims 1-3, 6-12, 14-19, and 21-23 merely disclose steps/components for performing matrix multiplication in [a] computer without further disclosing a practical application. Further, the claims appear to preempt every substantial practical application of the idea embodied by the claims.”

In response, Applicants point out that there would be no reason to preclude methods for improving efficiency of processing data on a computer as inherently being non-statutory subject matter, in view of the importance of computer in modern technology. Contrary to the Examiner’s characterization, Applicants respectfully submit that the present invention is not at all directed to preemption of a mathematical algorithm. Rather, it is directed to a specific method of presenting data for processing so that this processing becomes more efficient.

The cited references clearly demonstrate that efficient processing of this data is a well known problem in the art and that many, many ways are available for preparing the matrix data for this processing. The present inventors believe that their method provides an improvement over many of the methods currently used.

Therefore, contrary to the Examiner's characterization, Applicants believe that the method of the present invention is not at all a preemption of the mathematical algorithm itself for level 3 matrix multiplication. Rather, the method of the claimed invention provides a novel alternative to current method of preliminarily presenting the data for processing, thereby improving efficiency of the overall processing. The underlying mathematical algorithm of matrix multiplication is not in any way being preempted by this preliminary processing.

Stated slightly differently, the method of the present invention provides an improvement of the efficiency of the processing of the matrix data on a computer. As such, it provides a useful, concrete and tangible result and, therefore, fully satisfies the requirements for statutory subject matter for computer-implemented methods.

Contrary to the Examiner's characterization that the present preempts a mathematical algorithm, Applicants submit that a user will not infringe the present invention by simply processing the level 3 mathematical algorithm using any of the alternative methods for preliminarily handling of the data as described in any of the cited references. Therefore, there is no preemption of the mathematical algorithm itself.

The steps of the method of the claimed invention are related to the preparation and presentation of the data for processing and are not reasonably directed to claiming the mathematical algorithm *per se*.

Finally, it is noted that the Examiner's own rationale in the rejection for claims 21-23 for modifying primary reference Vinod by secondary reference Myszewski (e.g., "... *because it would enable to reduce computation stalling by optimizing the instruction*") clearly demonstrates the practical result provided by the present invention.

Relative to the remainder of the rejection for statutory subject matter, the Examiner additionally states:

"In addition, claims 14-19 and 22 [are] direct[ed] to a signal medium as clearly addressed in the specification page 26, particularly claim 19 clearly defines the signal bearing medium in the claim."

In response, Applicants respectfully submit that the description on page 26 clearly refers to the storage of instructions. There is no suggestion in this description that carrier waves or signals are being used for such storage and Applicants are not aware of any current technology that permits carrier waves or signals to be used for such storage of

instruction. Therefore, Applicants submit that this aspect of the rejection is based upon an incorrect interpretation of this description in the specification, wherein the Examiner attempts to take words in isolation and removes these words from the context of the sentence in the specification.

Moreover, independent claim 14 clearly states that the instructions be “tangibly embodied”, thereby precluding the interpretation that carrier waves or signals are involved, since neither carrier waves nor signals are considered in the art as “tangibly embodying” a set of instruction.

Therefore, Applicants again submit that all pending claims are clearly directed toward statutory subject matter and the Examiner is again requested to reconsider and withdraw this rejection.

III. THE PRIOR ART REJECTIONS

The Examiner alleges that newly-cited Vinod teaches the present invention described by 1-3, 6-12, and 14-18, and, when modified by Philip, renders obvious claim 19, and when modified by Myszewski, renders obvious claims 21-23.

Applicants respectfully disagree.

Newly-cited Vinod does not reasonably teach or suggest determining which of the three matrices will reside in cache, based upon size. Nor is there any suggestion to then stream the data from higher levels of cache for the remaining two matrices.

The Examiner points to Figure 2, § 3.2 on pages 9-10, and § 4.2 on pages 14-16 of Vinod as satisfying the independent claim limitation “... determining, based on sizes, for a level 3 matrix multiplication processing, which matrix will have data for a submatrix block residing in a lower level cache of said computer and which two matrices will have data for submatrix blocks residing in at least one higher level cache or a memory.”

In response, Applicants respectfully traverse that any of these sections or this figure reasonably has anything whatsoever to do with a discussion of relative size or even reasonably describes that only a selected one of three matrices will reside in L1 cache.

Moreover, in contrast to the discussion in Vinod, the present invention is actually only one part of the overall larger improvement package described in the seven co-pending applications identified at the beginning of the present application and is oriented toward SIMD machines not available or used at the time of Vinod. Of particular significance is

that, relative to Figure 2 on page 11 of Vinod, the present invention would be using a new data structure called register block format (see co-pending application S/N 10/671,888) rather than the row major data blocks shown in Figure 2, and which new data structure is designed to prevent data stalling that occurs with row major data in level 3 multiplication.

For the final independent claim limitation "... streaming data from said selected two matrices, for executing said level 3 matrix multiplication processing, so that said submatrix block residing in said lower level cache remains resident in said lower level cache", the Examiner points to §1.1 on pages 2-3, pages 5-6, and §3.2 on pages 9-10 of Vinod.

In response, Applicants respectfully traverse the Examiner's characterization that, to one having ordinary skill in the art, these locations reasonably describe the concept of streaming a selected two of three matrices from higher levels of cache while retaining a selected matrix as resident in L1 cache.

The only mention of "streaming" in Vinod occurs on the top of page 6 ("*The bottom tier operates on blocks that are assumed to be resident in L1 cache, streaming data into the CPU, keeping all the execution units in the processor as busy as possible and making maximum use of all the available processor features.*"), and describes a different context of streaming data, as meaning the data stream into the CPU from the L1 cache.

In contrast, the "streaming" of the claimed invention involves streaming data from higher levels of cache through L1 and into the CPU for two matrices (see dependent claim 6), with one matrix having data considered as resident in L1. This concept of streaming is entirely different from the simple streaming between the L1 cache and the CPU described in Vinod.

Moreover, in the present invention, there are only three streams of data, one for each matrix, with one considered as resident in L1 and the other two matrices are streamed through L1. In contrast, assuming the blocks shown in Figure 2 of Vinod have size $m_b \times m_b$ and $n_b \times n_b$, the number of streams that would be required, assuming that streaming were to be done similarly to that of the present invention, would be $2m_b + n_b$. Typical value for m_b and n_b is 4, so there would be 12 streams of data in Vinod, if it were to be converted into a similar streaming configuration. Again, the present invention achieves this minimal number of streams because of its use of more than one kernel and because it uses new data structures for the data that prevent the data stalling that would occur in Vinod with row major data blocks.

Applicants respectfully submit that the most that can reasonably be said about newly-cited Vinod is that it clearly demonstrates that presentation of the data for efficient level 3 matrix multiplication processing is clearly a known problem in the art and that there are many approaches considered in the art that address this problem.

Vinod provides a method that would appear to treat all three matrices equally in importance and that relies on many algorithms. In contrast, the present invention provides a method of presenting the data so that one matrix is selected to reside in the lower cache and the two remaining matrices are streamed from higher levels and uses only one algorithm. Vinod makes no suggestion of this technique of treating the three matrices differently for the presentation of the data, let alone the method of selecting the one matrix based upon size. Therefore, the Examiner is requested to place on record specific wording in Vinod that supports the rejection.

The Examiner does not rely upon secondary references Philip and Myszewski for overcoming this fundamental deficiency in Vinod, and these two references clearly do not provide a remedy for this deficiency.

Relative to the rejection for claim 19, based upon Philip as secondary reference, Applicants submit that this secondary reference fails to overcome the fundamental deficiency identified above for primary reference Vinod and only serves to clearly demonstrate that the present invention does indeed address a problem well known in the art and does indeed provide a result that is useful, concrete and tangible by its method of presenting data for processing.

Relative to the rejection for claims 21-23, based upon Myszewski as secondary reference, Applicants again submit that this secondary reference also serves to clearly demonstrate the well known problem of efficiency in level 3 matrix multiplication. Secondary reference Myszewski likewise does not overcome the fundamental deficiency of Vinod.

Moreover, Applicants respectfully traverse the Examiner's characterization that Figure 5 of Myszewski, or the cited locations (e.g., lines 55-64 of column 4, lines 34-55 of column 14, and columns 15-18) reasonably show anything related to selection of six possible alternate kernels or switching back and forth between two selected kernels.

Therefore, the Examiner is requested to cite specific wording that is relied upon as support for this allegation.

Finally, it is noted that the Examiner's rationale for modifying primary reference Vinod by secondary reference Myszewski (*e.g.*, "... *because it would enable to reduce computation stalling by optimizing the instruction*") clearly demonstrates the statutory subject matter of the present invention. Moreover, Applicants submit that this rationale fails to provide any indication that the urged modification of Vinod would reasonably provide any improvement over the method of primary reference Vinod, since Vinod already describes itself as providing this benefit.

In essence, Applicants submit that Vinod and Myszewski both clearly demonstrate that there are many methods available for preliminary processing of matrix data for level 3 matrix multiplication, all of which allege to overcome the same problem of computation inefficiency well known in the art. The present invention clearly provides an alternative to these methods that is clearly non-obvious.

IV. FORMAL MATTERS AND CONCLUSION

In view of the foregoing, Applicant submits that claims 1-3, 6-12, 14-19, and 21-23, all the claims presently pending in the application, are patentably distinct over the prior art of record and are in condition for allowance. The Examiner is respectfully requested to pass the above application to issue at the earliest possible time.

Should the Examiner find the application to be other than in condition for allowance, the Examiner is requested to contact the undersigned at the local telephone number listed below to discuss any other changes deemed necessary in a telephonic or personal interview.

The Commissioner is hereby authorized to charge any deficiency in fees or to credit any overpayment in fees to Assignee's Deposit Account No. 50-0510.

Respectfully Submitted,



Date: June 18, 2008

Frederick E. Cooperrider
Registration No. 36,769

McGinn Intellectual Property Law Group, PLLC
8321 Old Courthouse Road, Suite 200
Vienna, VA 22182-3817
(703) 761-4100
Customer No. 21254